Multiresolutional encoding and decoding in embedded image and video coders

Zixiang Xiong Beong-Jo Kim Pearlman, W.A.

Dept. of Electr. Eng., Hawaii Univ., Honolulu, HI ;

This paper appears in: **Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of the 1998 IEEE International Conference on**

Meeting Date: 05/12/1998 -05/15/1998

Publication Date: 12-15 May 1998

Location: Seattle, WA , USA

On page(s): 3709-3712 vol.6

Volume: 6, References Cited: 4

Number of Pages: 6 vol. lxiii+3816

INSPEC Accession Number: 6069265

Abstract:

We address multiresolutional encoding and decoding within the embedded zerotree wavelet (EZW) framework for both images and video. By varying a resolution parameter, one can obtain decoded images at different resolutions from one single encoded bitstream, which is already rate scalable for EZW coders. Similarly one can decode video sequences at different rates and different spatial and temporal resolutions from one bitstream. Furthermore, a layered bitstream can be generated with multiresolutional encoding, from which the higher resolution layers can be used to increase the spatial/temporal resolution of the images/video obtained from the low resolution layer. In other words, we have achieved full scalability in rate and partial scalability in space and time. This added spatial/temporal scalability is significant for emerging multimedia applications such as fast decoding, image/video database browsing, telemedicine, multipoint video conferencing, and distance learning

Index Terms:

decoding image coding image resolution image sequences transform coding video coding wavelet transforms EZW coders distance learning embedded image coders embedded video coders embedded zerotree wavelet image/video database browsing layered bitstream multimedia applications multipoint video conferencing multiresolution decoding multiresolution encoding rate scalability resolution parameter single encoded bitstream space scalability spatial resolution telemedicine temporal resolution time scalability video sequences

Documents that cite this document

MULTIRESOLUTIONAL ENCODING AND DECODING IN EMBEDDED IMAGE AND VIDEO CODERS

Zixiang Xiong[†], *Beong-Jo Kim*[‡], *and William A. Pearlman*[‡]
[†]Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822
[‡]Center for Image Processing Research, Rensselaer Polytechnic Institute, Troy, NY 12180

ABSTRACT

We address multiresolutional encoding and decoding within the embedded zerotree wavelet (EZW) framework for both images and video. By varying a resolution parameter, one can obtain decoded images at different resolutions from one single encoded bitstream, which is already rate scalable for EZW coders. Similarly one can decode video sequences at different rates and different spatial and temporal resolutions from one bitstream. Furthermore, a layered bitstream can be generated with multiresolutional encoding, from which the higher resolution layers can be used to increase the spatial/temporal resolution of the images/video obtained from the low resolution layer. In other words, we have achieved full scalability in rate and partial scalability in space and time. This added spatial/temporal scalability is significant for emerging multimedia applications such as fast decoding, image/video database browsing, telemedicine, multipoint video conferencing, and distance learning.

1. INTRODUCTION

With the fast evolution of multimedia systems, image and video compression is becoming the key enabling technology for delivering various image/video services over heterogeneous networks. Shapiro successfully developed a good practical EZW image coder [1], which was later improved by Said and Pearlman in their work of set partitioning in hierarchical trees (SPIHT) [2]. The SPIHT scheme was recently extended from 2-D to 3-D for video coding [3]. Although the coders in [1, 2, 3] offer scalability in rate, it is highly desirable to have temporal and/or spatial scalabilities in these coders for many applications such as video browsing and multicast network distributions.

We address both encoder and decoder spatial/temporal scalabilities within the EZW framework. Our presentation is based on the SPIHT image coder [2] as it is an improved version of Shapiro's EZW coder [1]. Unless otherwise specified, we henceforth assume that we are dealing with images instead of video. Like Shapiro's EZW coder, the SPIHT image coder uses zerotree quantization to efficiently predict the children nodes based on the significance/insignificance of their parent. It refines each wavelet coefficient on a bitmap base by successive set partitioning, and it stops when the size of the encoded bitstream reaches the exact target bitrate. The final encoded bitstream consists of a small header, sorting bits, sign bits, and refinement bits.

Since the SPIHT image coder in [2] is based on the multiresolutional wavelet decomposition, it is relatively easy to add multiresolutional encoding and decoding as functionalities in partial spatial/temporal scalability. We first concentrate on the simpler case of multiresolutional decoding, in which an encoded bitstream is assumed to be available at the decoder, and no modification to the encoder is needed. This approach is quite attractive since we do not need to change the encoder structure, and the decoder will work independently of the encoder. The idea of multiresolutional decoding is very simple: we partition the embedded bitstream into portions according to subbands, and only decode those that contribute to the resolution we want.

We then turn to multiresolutional encoding, where we modify the encoder such that the resulting multiresolutional encoder generates a layered bitstream, from which the higher resolution layers can be used to increase the spatial resolution of the image obtained from the low resolution layer. But modifying an embedded bitstream is not trivial, especially when the bitstream is generated using arithmetic coding (AC) whose efficiency depends on the context of the input source. Rearranging the SPIHT bitstream into layers before AC usually changes the final bitrate. Fortunately different bits in SPIHT are entropy coded using different modes. For example, the sorting bits are coded using several adaptive models in AC, whereas the sign bits and refinement bits are coded without modeling (i.e., by assuming uniform distribution). Thus we can keep the order of the sorting bits and rearrange the sign bits and refinement bits without changing the final bitrate after AC. This makes multiresolutional encoding possible as we can order the original bitstream into layers, with each layer corresponding to a different resolution (or portion). Although the layered bitstream is not fully embedded, the first layer is still rate scalable.

2. MULTIRESOLUTIONAL DECODING

In order to achieve multiresolutional decoding, we have to partition the encoded bitstream into portions according to subbands. This is done by putting flags in the bitstream during the process of decoding, when we scan through the bitstream and mark the portion that corresponds to the spatial locations defined by the decoder's desired resolution. As the encoded bitstream is embedded, this partitioning process can terminate at any point that is specified by the decoding bitrate. Fig. 1 shows such a bitstream partitioning for half resolution decoding. The gray portion of the bitstream contributes to the half resolution image, while the shaded portion corresponds to coefficients in the three highest frequency bands. We only decode the gray portion of the bitstream for the half resolution image. We also scale down the decoded wavelet coefficients by a power of two before applying the inverse wavelet transform. For decoding in lower resolutions, the gray portion of the bitstream in Fig. 1 is further partitioned in a similar manner.

0-7803-4428-6/98 $10.00 © 1998 IEEE

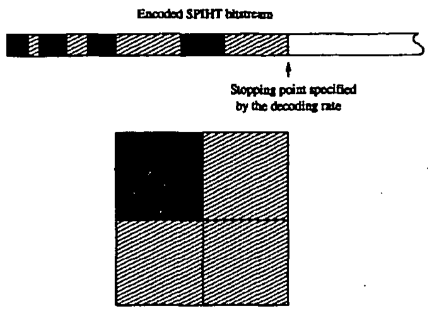


Figure 1: Partitioning of the encoded bitstream into portions according to subbands.

For the video case, we work on the encoded bitstream generated by the 3D SPIHT video coder [3], which is a direct generalization of the image coder [2]. There are two versions of the 3D SPIHT video coder: one with motion compensation (MC), one without. We use the version without MC as it is computationally much faster than the one with MC, and it gives comparable PSNR results to the H.263 video coder [4]. The basic bitstream partitioning method is the same as in the image case, except that we now have two types of flags: one for spatial scalability, another one for temporal scalability. We also scale down the 3D wavelet coefficients properly before applying the inverse wavelet transform.

3. MULTIRESOLUTIONAL ENCODING

The aim of multiresolutional encoding is to generate a layered bitstream. Although information bits (e.g., signs bits and refinement bits) corresponding to different resolutions in the original bitstream are interleaved, the SPIHT algorithm allows us to keep track of the spatial resolutions associated with these bits. Thus we can change the original encoder so that the new encoded bitstream is layered in spatial resolutions. Specifically, multiresolutional encoding amounts to putting in the first (low resolution) layer all the bits needed to decode a low resolution image, in the second (higher resolution) layer those to be added to the first layer for decoding a higher resolution image, and so on. This process is illustrated in Fig. 2 for the two-layer case, where scattered segments of the gray (and shaded) portion in the original bitstream are put together in the first (and second) layer of the new bitstream. A half resolution image can be decoded from the first layer (gray portion) alone, and a full resolution image from both the first and the second layers.

As the layered bitstream is a reordered version of the original embedded SPIHT bitstream, we lose overall scalability in rate after multiresolutional encoding. But the first layer (i.e., the gray layer in Fig. 2) of the layered bitstream is still embedded.

This multiresolutional encoding scheme for images can be readily extended to video in the 3D SPIHT coder. We also note that it is possible to combine multiresolutional encoding and decoding. For example, the first layer (gray portion) of the bitstream generated by the multiresolutional encoder can be used for decoding in lower resolutions.

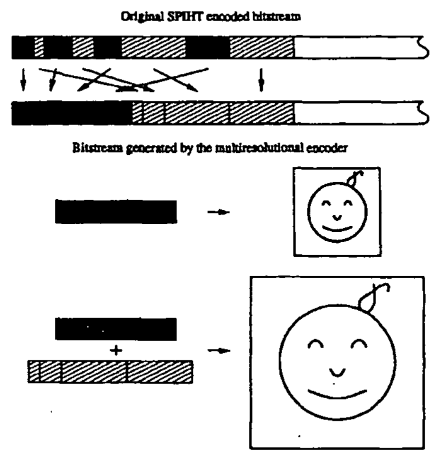


Figure 2: A multiresolutional encoder generates a layered bitstream, from which the higher resolution layers can be use to increase the spatial resolution of the image obtained from the low resolution layer.

4. RESULTS

We have implemented multiresolutional encoding and decoding for both the SPIHT image and video coders. For the image coder, versions with and without AC are both considered. Software for the image case, which is copyrighted, is available via anonymous ftp to `lena.eng.hawaii.edu` with the path `pub/scalability` in `encode_decode.tar.gz`.

4.1. MULTIRESOLUTIONAL DECODING

For results reported in this subsection, we assume a bitstream generated by the original encoder is available at the decoder. Fig. 5 shows the decoded Lena images at different resolutions and different rates from the same encoded bitstream generated by the SPIHT coder with AC. The full resolution 512×512 Lena image is decoded using six levels of inverse wavelet transform. We see that although the decoded full resolution images at relatively high (1 b/p) and very low (0.1 b/p) bitrates look quite different, this difference is much less pronounced between their corresponding half (and quarter) resolution decoded images. What this means is that we can set a very small (0.08 to 0.1 b/p) bitrate for fast decoding and browsing applications, saving decoding time.

Fig. 6 shows the first frames of decoded Foreman sequences at different spatial/temporal resolutions and different bitrates from the same 3D SPIHT encoded video bitstream. The full resolution QCIF Foreman sequence is decoded using three levels of inverse wavelet transform. The half resolution sequences are downsampled in both time and space.

In order to quantitatively assess the time savings in decoding multiresolutional images and video sequences, we give in Tables 1-2 the decoder running time (inverse wavelet transform and I/O included) on a SUN SPARC 20 for the

experiments reported in Figs. 5-6. Decoder running time without AC is also included for the image case. Results in Table 1 indicate that, without AC, half resolution image decoding is 3-4 times faster, quarter resolution image decoding can be 4-12 times faster than full resolution decoding. The time savings is smaller when AC is used, but quarter resolution image decoding can still be 4 times faster than full resolution decoding at 0.1 b/p. From Table 2, we see that, for video sequences, half resolution decoding can be as many as 4-5 times faster than full resolution decoding.

	Without AC		With AC	
	1 b/p	0.1 b/p	1 b/p	0.1 b/p
Full	2.95 sec	2.23 sec	4.90 sec	2.33 sec
Half	1.18 sec	0.58 sec	3.38 sec	0.78 sec
Quarter	0.75 sec	0.18 sec	2.98 sec	0.43 sec

Table 1: Decoding time for the Lena image at different rates and resolutions.

	27.59 kbits/s	20.97 kbits/s
Full (96 frames)	30.37sec	29.04 sec
Half (48 frames)	8.36 sec	6.61 sec

Table 2: Decoding time for the Foreman sequence at different rates and spatial resolutions.

endtable

4.2. MULTIRESOLUTIONAL ENCODING

Fig. 3 (a) shows the two-layer bitstream after multiresolutional encoding for Lena at 1 b/p with AC. Layer boundaries are marked in number of bytes. The decoded half and full resolution images are the same as those in Fig. 5 (a). That is: the first layer will be decoded to the half resolution image, and the first and the second layers together to the full resolution image in Fig. 5 (a), respectively. The difference between decoding from a layered bitstream and decoding from the original bitstream is that in the first case the decoder does not need to search for the whole bitstream for decoding lower resolution images. Fig. 3 (b) shows the two-layer bitstream after multiresolutional encoding for Lena at 0.1 b/p. The decoded half and full resolution images are the same as those in Fig. 5 (b). Table 3 gives the layer boundaries of the six-layer bitstreams for Lena encoded at 1 b/p and 0.1 b/p with AC.

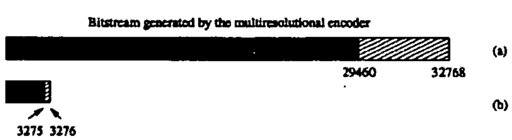


Figure 3: Two-layer bitstreams generated by multiresolutional encoding for Lena (layer boundaries marked in number of bytes). (a) Encoding at 1 b/p. (b) Encoding at 0.1 b/p.

Fig. 4 shows the layered bitstreams after multiresolutional 3D video encoding of the QCIF Foreman sequence. The decoded video sequences at different resolutions are the same as those in Fig. 6. The first layer will be decoded to the half resolution video, the first and the second layers to the full resolution video in Fig. 6, respectively.

Results in Figs. 3-4 indicate that, at relatively low bitrates, the second (or the last) layer of the bitstream is very short because most of the bits are used in coding the low resolution images/video. Thus multiresolutional encoding is more efficient at relatively high bitrate scenarios when a short first layer (or portion) can be used to decode the low resolution image/video. Also, as a large percentage of the bitstream is usually in the first layer, the time savings in low resolution decoding seems to result from the fewer levels of inverse wavelet transforms.

	1/64	1/32	1/16	1/8	1/4	Half	Full
1 b/p	22758	22893	23272	24198	26137	29460	32768
.1 b/p	2411	2478	2625	2865	3136	3275	3276

Table 3: Layer boundaries (in number of bytes) at different resolutions.

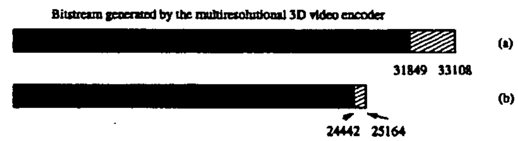


Figure 4: Layered bitstreams generated by multiresolutional 3D video encoding (layer boundaries marked in number of bytes). 96 frames of the QCIF Foreman sequence are encoded. (a) Encoding at a frame rate of 10 frames/sec and a bitrate of 27.59 kbits/sec. (b) Encoding at a frame rate of 10 frames/sec and a bitrate of 20.97 kbits/sec.

5. CONCLUSIONS

In this paper, we presented a simple way of decoding multiresolutional images and video sequences from a single embedded bitstream. An obvious advantage of it is the savings in decoding time. We also addressed multiresolutional encoding which has many applications in network communications. A good example will be scalable multicast video transmission in heterogeneous networks (ISDN, Internet, and Ethernets, etc.). Given the popularity of the SPIHT coder, we believe that our current work is a worthy addition to the original algorithm.

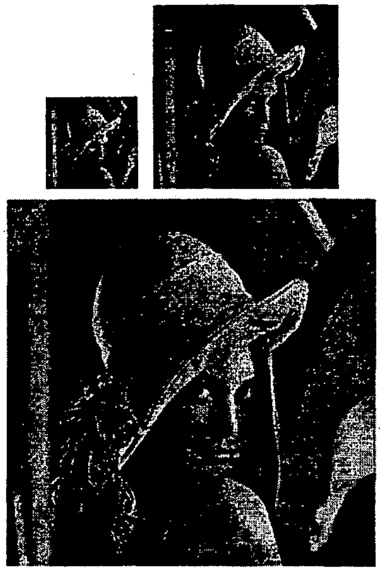
6. REFERENCES

[1] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol. 41, 12, pp. 3445-3463, December, 1993.

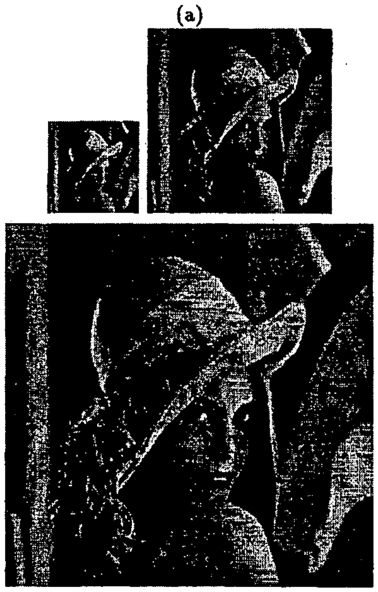
[2] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 6, 3, pp. 243-250, June, 1996.

[3] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," Proc. IEEE Data Compression Conference, pp. 251-260, 1997.

[4] ITU-T Recommendation H.263, "Video coding for low bitrate communication," Dec. 1995.

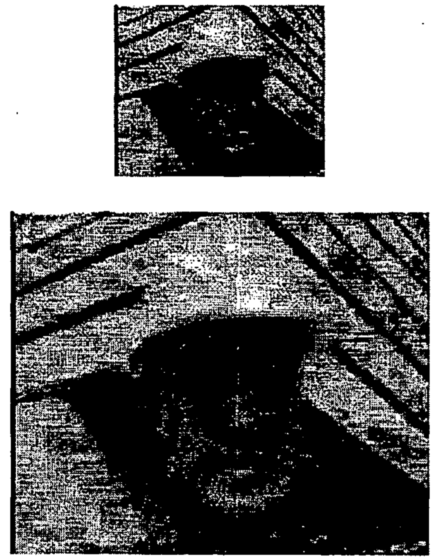


(a)

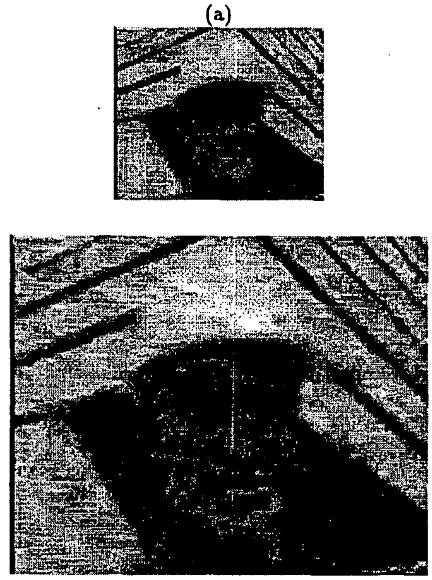


(b)

Figure 5: Multiresolutional decoding in the embedded SPIHT image coder [2] with arithmetic coding. (a) Decoded quarter, half, and full (512 × 512) resolution Lena images from the same encoded bitstream with the bitrate being set at 1 b/p. The full resolution image has a PSNR of 40.40 dB. (b) Decoded quarter, half, and full (512 × 512) resolution Lena images from the same encoded bitstream with the bitrate being set at 0.1 b/p. The full resolution image has a PSNR of 30.22 dB.



(a)



(b)

Figure 6: Multiresolutional decoding in the embedded 3D SPIHT video coder [3]. (a) The first frame of the half (both in space and time) and full resolution decoded QCIF (176 × 144) Foreman sequence. For the full resolution, encoding/decoding frame rate is 10 frames/sec, decoding bitrate is 27.59 kbits/sec. The average PSNR_Y over 96 frames is 26.63 dB, (b) The first frame of the half and full (both in space and time) resolution decoded QCIF Foreman sequence from the same embedded bitstream but with a decoding bitrate of 20.97 kbits/sec. The average PSNR_Y over 96 frames is 25.85 dB.

Your search matched **10** of **1058483** documents.
A maximum of **500** results are displayed, **15** to a page, sorted by **Relevance Descending** order.

Refine This Search:
You may refine your search by editing the current search expression or enter new one in the text box.

(partial or partially) and (decoding or decode or decomp Search

Check to search within this result set

Results Key:
JNL = Journal or Magazine **CNF** = Conference **STD** = Standard

---

1 **A new, fast, and efficient image codec based on set partitioning in hierarchical trees**
Said, A.; Pearlman, W.A.;
Circuits and Systems for Video Technology, IEEE Transactions on , Volume: 6 , Issue: 3 , June 1996
Pages:243 - 250

[Abstract] [PDF Full-Text (1200 KB)] **IEEE JNL**

---

2 **Partial encryption of compressed images and videos**
Cheng, H.; Xiaobo Li;
Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Sigı Processing, IEEE Transactions on] , Volume: 48 , Issue: 8 , Aug. 2000
Pages:2439 - 2451

[Abstract] [PDF Full-Text (260 KB)] **IEEE JNL**

---

3 **3-D wavelet compression and progressive inverse wavelet synthesi rendering of concentric mosaic**
Lin Luo; Yunnan Wu; Jin Li; Ya-Qin Zhang;
Image Processing, IEEE Transactions on , Volume: 11 , Issue: 7 , July 2002
Pages:802 - 816

[Abstract] [PDF Full-Text (485 KB)] **IEEE JNL**

---

4 **Arbitrarily shaped video-object coding by wavelet**
Guiwei Xing; Jin Li; Shipeng Li; Ya-Qin Zhang;
Circuits and Systems for Video Technology, IEEE Transactions on , Volume: 11 , Issue: 10 , Oct. 2001
Pages:1135 - 1139

[Abstract] [PDF Full-Text (96 KB)] **IEEE JNL**

5 **Arbitrarily shaped video object coding by wavelet**
Guiwei Xing; Jin Li; Shipeng Li; Ya-Qin Zhang;
Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEE
International Symposium on , Volume: 3 , 28-31 May 2000
Pages:535 - 538 vol.3

[Abstract] [PDF Full-Text (384 KB)] **IEEE CNF**

6 **Similarity measures for compressed image databases**
Sangassapaviriya, P.; Ogunbona, P.O.;
TENCON '97. IEEE Region 10 Annual Conference. Speech and Image Technolc
for Computing and Telecommunications'., Proceedings of IEEE , Volume: 1 , 2
Dec. 1997
Pages:203 - 206 vol.1

[Abstract] [PDF Full-Text (384 KB)] **IEEE CNF**

7 **Data compression and encryption using cellular automata transforn**
Lafe, O.;
Intelligence and Systems, 1996., IEEE International Joint Symposia on , 4-5 I
1996
Pages:234 - 241

[Abstract] [PDF Full-Text (520 KB)] **IEEE CNF**

8 **Reduced memory requirements in modified JPEG2000 codec**
Lafruit, G.; Andreopoulos, Y.; Masschelein, B.;
Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conferen
on , Volume: 1 , 1-3 July 2002
Pages:219 - 226 vol.1

[Abstract] [PDF Full-Text (973 KB)] **IEEE CNF**

9 **Multiresolutional encoding and decoding in embedded image and vi coders**
Zixiang Xiong; Beong-Jo Kim; Pearlman, W.A.;
Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of t
1998 IEEE International Conference on , Volume: 6 , 12-15 May 1998
Pages:3709 - 3712 vol.6

[Abstract] [PDF Full-Text (652 KB)] **IEEE CNF**

10 **Embedded zerotree based lossless image coding**
Egger, O.; Kunt, M.;
Image Processing, 1995. Proceedings., International Conference on , Volume
3 , 23-26 Oct. 1995
Pages:616 - 619 vol.3

[Abstract] [PDF Full-Text (724 KB)] **IEEE CNF**

Patent Assignment Abstract of Title

Total Assignments: 1

Application #: 09579803 **Filing Dt:** 05/26/2000 **Patent #:** NONE **Issue Dt:**

PCT #: NONE **Publication #:** NONE **Pub Dt:**

Inventors: Takahiro Fukuhara, Seiji Kimura, Hitoshi Kiya

Title: Wavelet inverse transform method and apparatus and wavelet decoding method and apparatus

Assignment: 1

Reel/Frame:	**Received:**	**Recorded:**	**Mailed:**	**Pages:**
011139/0981	10/12/2000	10/02/2000	12/11/2000	2

Conveyance: ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).

Assignors: FUKUHARA, TAKAHIRO **Exec Dt:** 07/21/2000

KIMURA, SEIJI **Exec Dt:** 07/24/2000

KIYA, HITOSHI **Exec Dt:** 09/20/2000

Assignees: SONY CORPORATION
7-35 KITASHINAGAWA 6-CHOME SHINAGAWA-KU
TOKYO 141-0001, JAPAN

KIYA, HITOSHI
5-9-3-307, MINAMI-OHSAWA, HACHIOJI-SHI
TOKYO, JAPAN

Correspondent: FROMMER, LAWRENCE & HAUG LLP
WILLIAM S. FROMMER
745 FIFTH AVENUE
NEW YORK N.Y. 10151

Search Results as of: 2/22/2004 11:35:16 A.M.

US-PAT-NO: 6301370

DOCUMENT-IDENTIFIER: US 6301370 B1
See image for Certificate of Correction

TITLE: Face recognition from video images

DATE-ISSUED: October 9, 2001

US-CL-CURRENT: 382/103, 342/90 , 348/169 , 382/117 , 382/118 , 382/190 , 382/209 , 382/276

APPL-NO: 09/ 206195

DATE FILED: December 4, 1998

PARENT-CASE:

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims priority under 35 U.S.C. .sctn.119(e)(1) and 37 C.F.R. .sctn.1.78(a)(4) to U.S. provisional application serial number 60/081,615 filed Apr. 13, 1998 and titled VISION ARCHITECTURE TO DESCRIBE FEATURES OF PERSONS.

---------- KWIC ---------

Claims Text - CLTX (3):
transforming **only** the bound **portion** and not the

unbound **portion** of the image
frame using a **wavelet** transformation to generate a transformed image;

Claims Text - CLTX (38):

means for transforming **only** the bound **portion** and not the unbound **portion** of the image frame using a **wavelet** transformation to generate a transformed image;

Claims Text - CLTX (48):

means for transforming **only** the bound **portion** and not the unbound **portion** of each image frame using a **wavelet** transformation to generate a transformed image;

US-PAT-NO: 6021224

DOCUMENT-IDENTIFIER: US 6021224 A

TITLE: Multiresolution lossless/lossy compression and storage of data for efficient processing thereof

DATE-ISSUED: February 1, 2000

US-CL-CURRENT: 382/232

APPL-NO: 08/ 829457

DATE FILED: March 28, 1997

---------- KWIC ---------

Detailed Description Text - DETX (50):

Described in detail above is a technique for compressing data and storing the compressed data in such a manner that enables efficient retrieval and processing of selected portions of the compressed data. With the exception of the quantization and lossless encoding operations, all other operations can be performed linearly. This means that if the lossless encoding is performed on separate blocks of the image (and the residual) independently, a portion of the image of interest can be extracted without having to read and decode the entire image. In particular, a lossy version of the image

can be reconstructed at any
level of resolution by just decoding the blocks
that contain **wavelet**
coefficients corresponding to the required **portion,**
and inverting the **wavelet**
transform for these coefficients, **only**. This
provides a significant speedup
during the decoding process, since the whole image
does not need to be
processed, and allows image processing operations
to be efficiently applied to
reduced resolution image constructs.

Detailed Description Text - DETX (61):
C. Speed of retrieval of selected portions of
the information. Here,
selected **portions** of the information include not
only selected **portions** of the
image, but, for instance, in the case of the
wavelet transform, selected
portions of the different subbands.

US-PAT-NO: 6421463

DOCUMENT-IDENTIFIER: US 6421463 B1
See image for Certificate of Correction

TITLE: Trainable system to search for objects in images

DATE-ISSUED: July 16, 2002

US-CL-CURRENT: 382/224, 382/279

APPL-NO: 09/ 282742

DATE FILED: March 31, 1999

PARENT-CASE:

This application claims benefit of provisional appln. No. 60/080,358 filed Apr. 1, 1998.

---------- KWIC ---------

Abstract Text - ABTX (1):
A trainable object detection system and technique for detecting objects such as people in static or video images of cluttered scenes is described. The described system and technique can be used to detect highly non-rigid objects with a high degree of variability in size, shape, color, and texture. The system learns from examples and does not rely on any a priori (hand-crafted)

models or on motion. The technique utilizes a **wavelet** template that defines the shape of an **object in terms of a subset of the wavelet coefficients** of the image. It is invariant to changes in color and texture and can be used to robustly define a rich and complex class of objects such as people. The invariant properties and computational efficiency of the wavelet template make it an effective tool for object detection.

Brief Summary Text - BSTX (18):

In accordance with the present invention, an object detection system includes (a) an image preprocessor for moving a window across the image and a classifier coupled to the preprocessor for classifying the portion of the image within the window. The classifier includes a **wavelet** template generator which generates a **wavelet** template that defines the shape of an **object with a subset of the wavelet coefficients** of the image. The wavelet template generator generates a wavelet template which includes a set of regular regions of different scales that correspond to the support of a subset of significant wavelet functions. The relationships between different regions are expressed as constraints on the values of the wavelet coefficients. With this particular arrangement, a system which is trainable and which detects objects in static or video images of cluttered scenes is provided. The wavelet template defines an object as a set of regions and relationships among the regions. Use of a

wavelet basis to represent the template yields both a computationally efficient technique and an effective learning scheme. By using a **wavelet** template that defines the shape of an **object in terms of a subset of the wavelet coefficients** of the image, the system can detect highly non-rigid objects such as people and other objects with a high degree of variability in size, shape, color, and texture. The wavelet template is invariant to changes in color and texture and can be used to robustly define a rich and complex class of objects such as people. The system utilizes a model that is automatically learned from examples and thus can avoid the use of motion and explicit image segmentation to detect objects in an image. The system further includes a training system coupled to the classifier and including a database including both positive and negative examples; and a quadratic programming solver. The system utilizes a general paradigm for object detection. The system is trainable and utilizes example-based models. Furthermore, the system is reconfigurable and extendible to a wide variety of object classes.

Detailed Description Text - DETX (40):

As can be observed in FIG. 2B, there are no consistent patterns in the color and texture of pedestrians or their backgrounds in arbitrary cluttered scenes in unconstrained environments. This lack of clearly discernible interior features is circumvented by relying on (1) differences in the intensity between

pedestrian bodies and their backgrounds and (2) consistencies within regions inside the body boundaries. The **wavelet coefficients can be interpreted as indicating an almost uniform area,** i.e. "no-change", if their absolute value is relatively small, or as indicating "strong change" if their absolute value is relatively large. The wavelet template sought to be identified consists solely of wavelet coefficients (either vertical, horizontal or corner) whose types ("change"/"no-change") are both clearly identified and consistent along the ensemble of pedestrian images; these comprise the "important" coefficients.

Detailed Description Text - DETX (63):

Once the important basis functions have been identified, various classification techniques can be used to learn the relationships between the **wavelet coefficients that define the object** class. The system can be trained using the bootstrapping technique described above in conjunction with FIG. 2.

US-PAT-NO: 6501859

DOCUMENT-IDENTIFIER: US 6501859 B1

TITLE: Image compression using wavelet data or original image data depending on code amount

DATE-ISSUED: December 31, 2002

US-CL-CURRENT: 382/239, 382/240

APPL-NO: 09/ 370198

DATE FILED: August 9, 1999

FOREIGN-APPL-PRIORITY-DATA:

COUNTRY	APPL-NO	APPL-DATE
JP	10-237074	August 24, 1998

---------- KWIC ---------

Detailed Description Text - DETX (78):

Note that the size of an image, the number of bits per pixel, and the like, which are required for decoding are appropriately appended to the generated code sequence. In this embodiment, since appropriate **wavelet** transform level counts are selected in units of blocks, highly efficient encoding can be done in correspondence with the local natures of an

image, and since an image is
divided in units of blocks, **only a portion** that the
user wants to see can be
quickly decoded.

PAT-NO: JP02001333430A

DOCUMENT-IDENTIFIER: JP 2001333430 A

TITLE: IMAGE PROCESSING UNIT, METHOD, AND COMPUTER-READABLE STORAGE MEDIUM

PUBN-DATE: November 30, 2001

INVENTOR-INFORMATION:

NAME COUNTRY

HATORI, KENJI N/A

ASSIGNEE-INFORMATION:

NAME COUNTRY

CANON INC N/A

APPL-NO: JP2000151382

APPL-DATE: May 23, 2000

INT-CL (IPC): H04N007/30, H03M007/30 , H04N001/41 , H04N005/232 , H04N005/92

ABSTRACT:

PROBLEM TO BE SOLVED: To provide an image processing unit that can reproduce a moving picture and only part of consecutively shot still pictures and

reproduce and display the image through magnification without the need for hardware such as a high-speed CPU.

SOLUTION: In the case of expanding **only** a designated **area** by an instruction of a user, a system inverse quantization section 8 expands **only an area** designated by an **area** instruction section 11 at coding, and e.g. an **inverse discrete wavelet** transform section 9 substitutes a white or black pixel for an image signal in an **area** not decoded to allow **only** a star-shaped **area** part designated at coding to be decoded and to allow the other parts to be displayed as a white or black level, or substitutes an invalid value for the **area** not decoded as an image signal and segments and displays the part including the designated **area**.

US-PAT-NO: 6356666

DOCUMENT-IDENTIFIER: US 6356666 B1

TITLE: Image compressing/expanding method and image compression expanding device

DATE-ISSUED: March 12, 2002

US-CL-CURRENT: 382/240, 382/232 , 382/239

APPL-NO: 09/ 242538

DATE FILED: February 18, 1999

PARENT-CASE:

This application is the national phase under 35 U.S.C. .sctn.371 of prior PCT International Application No. PCT/JP97/02655 which has an International filing date of Jul. 30, 1997 which designated the United States of America.

PCT-DATA:
APPL-NO: PCT/JP97/02655
DATE-FILED: July 30, 1997
PUB-NO: WO99/07155
PUB-DATE: Feb 11, 1999
371-DATE: Feb 18, 1999
102(E)-DATE: Feb 18, 1999

---------- KWIC ---------

Brief Summary Text - BSTX (10):

Further, a problem arises in that since the number of transformation coefficients zero in value relatively increases if an **inverse lossless wavelet** transformation is effected on wavelet transformation **coefficients comprised of only** information necessary for updating, it is quite natural that the number of arithmetic operations at the **inverse lossless wavelet** transformation can be reduced and the processing can be speeded up by using its relative increase, but the number of the transformation coefficients zero in value relatively decreases because the **inverse lossless wavelet** transformation is made after the updating of information in a transformation coefficient domain. Thus the speeding up of the processing utilizing the property that many coefficients zero in value are included in the wavelet transformation coefficients indicative of the information necessary for updating cannot be expected. Namely, the image compressing/expanding device based on the linear transformation and the image compressing/expanding device based on the lossless wavelet transformation have both merits and demerits respectively.

Brief Summary Text - BSTX (27):

an image expanding device having at least one Inverse S transformation part for implementing an **inverse lossless wavelet transformation by an inverse** S transformation, and wherein the updated auxiliary information is generated only

from LSB of respective lossless wavelet S transformation coefficients necessary to generate a pre-updating image through a lossless **wavelet inverse** S transformation and LSB of lossless wavelet S transformation **coefficients bearing only** information necessary for updating.

Brief Summary Text - BSTX (31):

an image expanding device having at least one inverse TS transformation part (constructed by an S transformation and a linear transformation and a non-linear transformation placed in its previous stage) for implementing an **inverse lossless wavelet transformation by an inverse** TS transformation, and wherein the generation of updated auxiliary information under a progressive reproducing process through an inverse transformation set in a stage preceding an inverse S transformation is performed as a first stage based only on the two rightmost bits of respective transformation coefficients necessary to generate pre-updating data through the same inverse transformation and the two rightmost bits of respective transformation **coefficients bearing only** information newly necessary for updating, and as a second stage, the so-obtained updated information is progressively reconstructed through the inverse S transformation as an input, whereby the progressive reconstruction display of an image at the image expanding device is performed in accordance with the two-stage processing through the inverse S transformation.

Claims Text - CLTX (29):

an image expanding device having at least on Inverse S transformation part for implementing an **inverse lossless wavelet transformation by an inverse** S transformation, and wherein the image-domain update information is generated only from LSB of respective lossless wavelet S transformation coefficients necessary to generate a pre-updating image through a lossless **wavelet inverse** S transformation and LSB of lossless wavelet S transformation **coefficients bearing only** information necessary for updating.

Claims Text - CLTX (32):

an image expanding device having at least one inverse TS transformation part for implementing an **inverse lossless wavelet transformation by an inverse** TS transformation, and wherein the generation of image-domain update information under a progressive reconstructing process through an inverse transformation set in a stage preceding an inverse S transformation is performed as a first stage based only on the two rightmost bits of respective transformation coefficients necessary to generate pre-updating data through the same inverse transformation and the two rightmost bits of respective transformation **coefficients bearing only** information newly necessary for updating, and as a second stage, the so-obtained update information is progressively reconstructed through the inverse S transformation as an input, whereby the progressive

reconstruction display of an image at the image expanding device is performed in accordance with the two-stage processing through the inverse S transformation.

Claims Text - CLTX (60):

at least on Inverse S transformation part for implementing an **inverse lossless wavelet transformation by an inverse** S transformation, and wherein the image-domain update information is generated only from LSB of respective lossless wavelet S transformation coefficients necessary to generate a pre-updating image through a lossless **wavelet inverse** S transformation and LSB of lossless wavelet S transformation **coefficients bearing only** information necessary for updating.

Claims Text - CLTX (62):

at least one inverse TS transformation part for implementing an **inverse lossless wavelet transformation by an inverse** TS transformation, and wherein the generation of image-domain update information under a progressive reconstructing process through an inverse transformation set in a stage preceding an inverse S transformation is performed as a first stage based only on the two rightmost bits of respective transformation coefficients necessary to generate pre-updating data through the same inverse transformation and the two rightmost bits of respective transformation **coefficients bearing only**

information newly necessary for updating, and as a second stage, the so-obtained update information is progressively reconstructed through the inverse S transformation as an input, whereby the progressive reconstruction display of an image at the image expanding device is performed in accordance with the two-stage processing through the inverse S transformation.